Marriott Statistical Consulting

Comparisons with a Critical Concentration

What's Changed in the New Guidance & Why

Nigel Marriott CStat

Independent Statistician

nigel@marriott-stats.com

On behalf of CL:AIRE at SILC Annual Forum, London, March 2020

www.marriott-stats.com

2 Temple Street, Keynsham, Bristol, BS31 1EG, United Kingdom
Registered in England, Company Number 5577275, VAT number GB883304029

- Brexit, Pexit & Clexit –

The backstory to how we got here today

The World Before 2016

BREXIT

Since the 1950's, there has been endless debates about the UK's relationship with Europe.

The relationship has changed over time following various negotiations and treaties.

The pros and cons of EU membership continued to be endlessly debated.

P:EXIT

Since the 1930's, when **Null Hypothesis Significance Testing** was formalised, statisticians have been aware of the many ways **P-values** can be misused.

Whilst much scientific research has benefited from P-values, issues such as **P-value Hacking** and the **Reproducibility Crisis** are of increasing concern.

CL:EXIT

The existing 2008 guidance is a textbook example of **1-way hypothesis testing** with P-values being used to decide if the requirements of the 1990 EPA have been met.

In practice, concerns were being raised that the guidance was not being used or was being used in "bright-line" fashion to make decisions.

In 2016 …

BREXIT

The UK votes 52:48 to leave the EU.

It is unclear what the next step is though and the UK enters a tumultuous period of politics.

P:EXIT

The ASA (**American Statistical Association)** publishes a definitive **Statement on P-Values** **https://bit.ly/2TwPd20**

The statement focuses on what P-values are not and how they should not be used.

It says very little about how statistical inference should be done instead except for one key paragraph (next slide)

CL:EXIT

CL:AIRE convenes a steering group to rewrite the 2008 guidance. This meets for the first time in Coventry in November 2016.

Initially it was agreed the focus on the **Statistical Guidance**, not the **Conceptual Site Model** or **Sampling Plans** …

Finally in 2019 …

BREXIT

After …
2 General Elections,
2 Prime Minsters,
3 Meaningful Votes,
2 missed deadlines,
resignations, court cases, parliamentary shenanigans galore …

… Parliament approves the Withdrawal Act in December 2019 paving the way for the UK's exit in Jan 2020.

P:EXIT

The ASA Statement was widely covered & discussed in scientific circles around the world.
https://bit.ly/38c2ORG

It initiated much debate on how scientific inference should be done instead and in March 2019, the ASA published "**Moving to a World Beyond P<0.05**"
https://bit.ly/2we28Ox

CL:EXIT

… but after the first draft was written, there was concern that existing guidance on CSMs and Sampling was inadequate. This led to much back and forth on how much the revised guidance should cover these as well.

The final version (of Dec 2019) hopefully is nearly agreed and will be published shortly.

2020 & Beyond - This is not the end ...

BREXIT

Current negotiations with the EU over trade is just the start.

Brexit involves swapping a known but flawed quantity for an uncertain future and resolving that **uncertainty** will take a number of years.

P:EXIT

The 2019 ASA Statement gives many ideas on how to interpret data instead of just using P-values. However, there is no one recommended alternative.

Pexit involves swapping the straightforward if flawed idea of P-values for an **uncertain & diverse** range of methods.

CL:EXIT

I was keen to ensure that the new guidance followed the principles of Pexit to deliver Clexit.

It means that CL:AIRE has swapped a guidance that could be described as prescriptive for one that places greater reliance on expert **judgement**. I recognise that this will create some uncertainty!

2016 ASA P-Value Statement – Key Paragraph

"Good statistical practice, as an essential component of good scientific practice, emphasizes …

1. *… principles of good study design and conduct,*
2. *… a variety of numerical and graphical summaries of data,*
3. *… understanding of the phenomenon under study,*
4. *… interpretation of results in context,*
5. *… complete reporting and*
6. *… proper logical and quantitative understanding of what data summaries mean. …*

… No single index should substitute for scientific reasoning."

2019 ASA P-Value Statement – Key Paragraphs

"Yet the voices in the 43 papers in this issue do not sing as one. At times in this editorial and the papers you'll hear deep dissonance, the echoes of "statistics wars" still simmering today (Mayo 2018). At other times you'll hear melodies wrapping in a rich counterpoint that may herald an increasingly harmonious new era of statistics. To us, these are all the sounds of statistical inference in the 21st century, the sounds of a world learning to venture beyond "p < 0.05."

This is a world where researchers are free to treat "p = 0.051" and "p = 0.049" as not being categorically different, where authors no longer find themselves constrained to selectively publish their results based on a single magic number. In this world, where studies with "p < 0.05" and studies with "p > 0.05" are not automatically in conflict, researchers will see their results more easily replicated—and, even when not, they will better understand why.

As we venture down this path, we will begin to see fewer false alarms, fewer overlooked discoveries, and the development of more customized statistical strategies. Researchers will be free to communicate all their findings in all their glorious uncertainty, knowing their work is to be judged by the quality and effective communication of their science, and not by their p-values. As "statistical significance" is used less, statistical thinking will be used more."

- The New Guidance –
What's changed and how to use it

APPENDIX A1 - "Good statistical practice, as an essential component of good scientific practice, emphasizes …

1. ***"… principles of good study design and conduct …"***
   - ❑ Section 2 of the New Guidance states the essential requirements that need to be taken care of before it can be used i.e.
   - ❑ **Sampling Plans, CSMs, Non-Detect rules, Critical Concentration**
   - ❑ New Guidance does not cover how to make decisions on these points.
   - ❑ Instead links to other documents are given.

2. *"… a variety of numerical and graphical summaries of data …"*
3. *"… understanding of the phenomenon under study …"*
4. *"… interpretation of results in context …"*
5. *"… complete reporting and …"*
6. *"… proper logical and quantitative understanding of what data summaries mean. …"*

… No single index should substitute for scientific reasoning."

APPENDIX A1 - "Good statistical practice, as an essential component of good scientific practice, emphasizes …

1. *"… principles of good study design and conduct …"*

2. ***"… a variety of numerical & graphical summaries of data …"***
   - ❑ Section 4 describes how to produce and calculate these outputs.
   - ❑ **Dot, Box & Spatial Plots, Summary Statistics & Confidence Intervals**
   - ❑ This list is NOT intended to be exhaustive and you are free to summarise your data with other tabular & chart formats.
   - ❑ Rationale for using 2-way Confidence Intervals instead of 1-way Confidence Intervals is given in Appendix A2.

3. *"… understanding of the phenomenon under study …"*
4. *"… interpretation of results in context …"*
5. *"… complete reporting and …"*
6. *"… proper logical and quantitative understanding of what data summaries mean. …"*

… No single index should substitute for scientific reasoning."

APPENDIX A1 - "Good statistical practice, as an essential component of good scientific practice, emphasizes …

1. *"… principles of good study design and conduct …"*
2. *"… a variety of numerical and graphical summaries of data …"*

3. ***"… understanding of the phenomenon under study …"***
   - ❑ Using the CSM to identify expected Data Type and results.
   - ❑ Appendix B describes 3 broad **Data Types** expected in real life.
   - ❑ Section 3 introduces 3 example data sets used in the guidance.
   - ❑ The concept of Data Types was introduced after much debate to help practitioners make a decision on whether the sample size is sufficient.
   - ❑ **It does NOT replace the requirement to design your sampling plan in accordance with statistical principles as per step 1 above!**

4. *"… interpretation of results in context …"*
5. *"… complete reporting and …"*
6. *"… proper logical and quantitative understanding of what data summaries mean. …"*

… No single index should substitute for scientific reasoning."

APPENDIX A1 - "Good statistical practice, as an essential component of good scientific practice, emphasizes …

1. *"… principles of good study design and conduct …"*
2. *"… a variety of numerical and graphical summaries of data …"*
3. *"… understanding of the phenomenon under study …"*

4. ***"… interpretation of results in context …"***
   - ❑ Are your results very different from what your CSM indicated?
   - ❑ Do you need to go back and revise your CSM first?
   - ❑ Given the expected and actual data type, do you need more samples?
   - ❑ Is the decision clear & obvious or borderline?
   - ❑ Section 5 uses the 3 example data sets to explore these questions.
   - ❑ **This is true Pexit, the use of scientific judgement SUPPLEMENTED by statistical results, not replaced by P-values.**

5. *"… complete reporting and …"*
6. *"… proper logical and quantitative understanding of what data summaries mean. …"*

… No single index should substitute for scientific reasoning."

APPENDIX A1 - "Good statistical practice, as an essential component of good scientific practice, emphasizes …

1. *"… principles of good study design and conduct …"*
2. *"… a variety of numerical and graphical summaries of data …"*
3. *"… understanding of the phenomenon under study …"*
4. *"… interpretation of results in context …"*

5. ***"… complete reporting and …"***
   - ❑ Section 6 lists what is needed for complete reporting.
   - ❑ It starts out by asking if you are in a position to state "***I am confident that …"*** for a variety of points.
   - ❑ It notes that your level of confidence may differ if you are in a Planning scenario compared to a Part 2A remediation scenario.

6. *"… proper logical and quantitative understanding of what data summaries mean. …"*

… No single index should substitute for scientific reasoning."

APPENDIX A1 - "Good statistical practice, as an essential component of good scientific practice, emphasizes …

1. *"… principles of good study design and conduct …"*
2. *"… a variety of numerical and graphical summaries of data …"*
3. *"… understanding of the phenomenon under study …"*
4. *"… interpretation of results in context …"*
5. *"… complete reporting and …"*

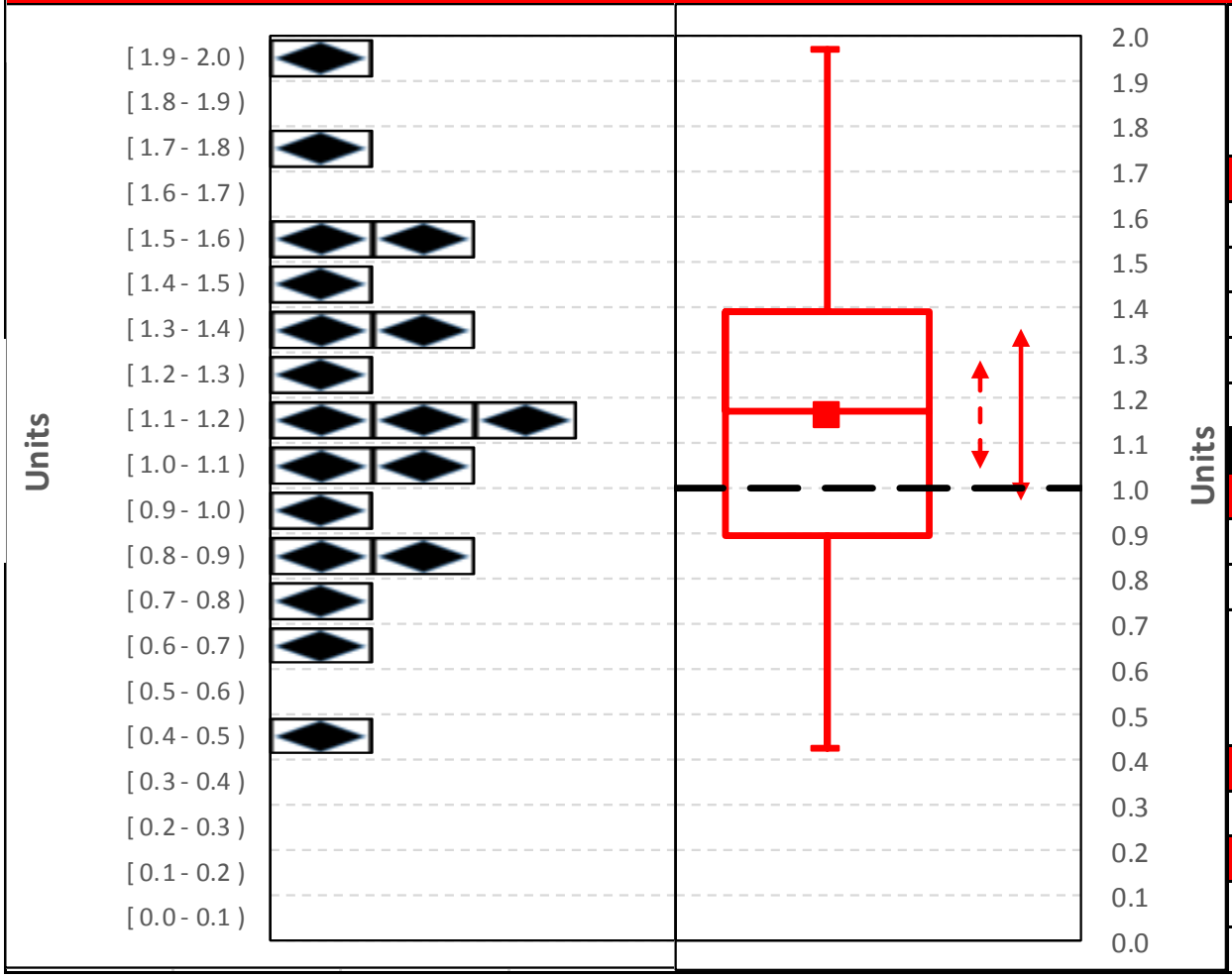
6. ***"… proper logical and quantitative understanding of what data summaries mean. …"***

- The new guidance is based on comparing confidence intervals to a critical concentration but this can only done once steps 1 to 5 have been undertaken.
- If the comparison indicates a borderline decision, what is the right and wrong way to interpret what you see?
- **The 3 example data sets explored in Section 5 are deliberately intended to be borderline examples** to illustrate how to use statistical reasoning and are shown in the next 3 slides.

… No single index should substitute for scientific reasoning."

DATASET A – Symmetric Data Type, Part 2A Scenario

Dot & Box Plots, Summary Statistics & Confidence Intervals - SYMMETRIC



Summary Statistics Raw Values

#Samples	19
Maximum	2.0
Upp Quartile	1.4
Median	1.2
Lwr Quartile	0.9
Minimum	0.4
Critical Conc	1.0
Mean	1.2
Std Deviation	0.4
IQR	0.5

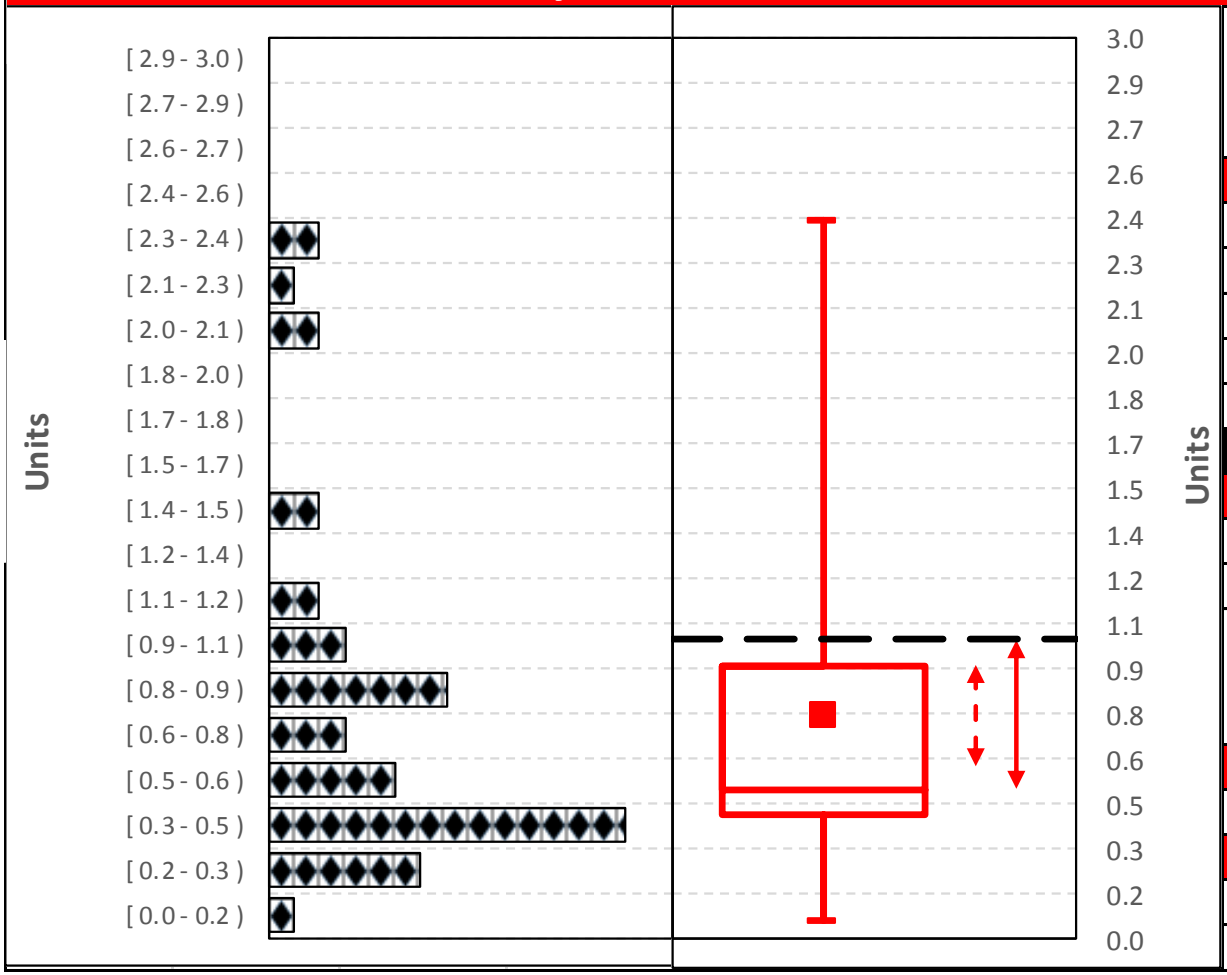
Confidence Intervals

80% Conf Interval	
1.0 ⟷	1.3
95% Conf Interval	
1.0 ⟷	1.4

Previously, H0: Mean >=1, H1: Mean < 1, P-Value = 4.3%

DATASET B – Log-Symmetric Data Type, Planning

Dot & Box Plots, Summary Statistics & Confidence Intervals - LOG SYMMETRIC



Summary Statistics Raw Values

#Samples	48
Maximum	2.4
Upp Quartile	0.9
Median	0.5
Lwr Quartile	0.4
Minimum	0.1
Critical Conc	1.0
Mean	0.7
Std Deviation	0.6
IQR	0.5

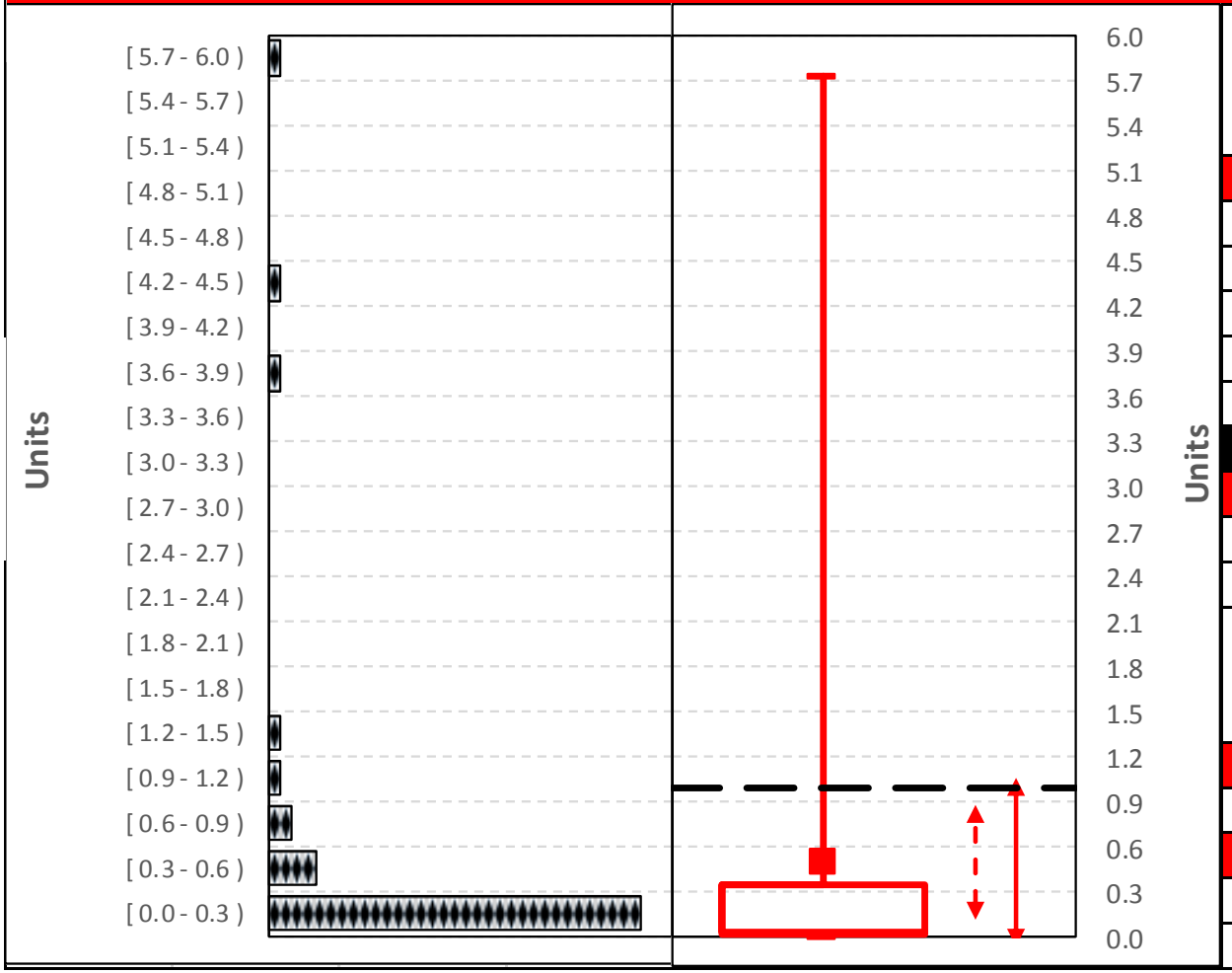
Confidence Intervals

95% Conf Interval	
0.6 ⟷	0.9
99.5% Conf Interval	
0.5 ⟷	1.0

Previously, H0: Mean < 1, H1: Mean >= 1, P-Value = 0.2%

DATASET C – Fat-Tailed Data Type, Planning

Dot & Box Plots, Summary Statistics & Confidence Intervals - FAT-TAILED



Summary Statistics Raw Values

#Samples	42
Maximum	5.7
Upp Quartile	0.4
Median	0.0
Lwr Quartile	0.0
Minimum	0.0
Critical Conc	1.0
Mean	0.5
Std Deviation	1.2
IQR	0.3

Confidence Intervals

95% Conf Interval	
0.1 ↔	0.9
99.5% Conf Interval	
-0.1 ↔	1.1

Previously, H0: Mean < 1, H1: Mean >= 1, P-Value = 0.5% or 14.1%

Appendix B - How to Identify Data Type

- Suppose two people are selected at random from the UK and you are told only the following information and nothing else.
  - ❑ The sum of their heights is 13 feet.
  - ❑ The sum of their net wealth is £20 million.
- Clearly these are exceptional individuals but which of the two scenarios is more likely for height?
  - A. One person is 2 feet tall and the other is 11 foot tall.
  - B. Both are 6 foot 6 inches.
- And which of the two scenarios is more likely for wealth?
  - A. One person has £20mn net and the other zero net wealth.
  - B. Both have £10mn net.
- I strongly suspect you have given different answers!
- This is because Heights & Wealth have different data types. The same goes for contamination data.
  - A. This is the **FAT-TAILED** data type.
  - B. This is the **SYMMETRIC** data type.
  - C. Unsure? Probably **LOG SYMMETRIC**.

Marriott Statistical Consulting

Comparisons with a Critical Concentration

What's Changed in the New Guidance & Why

Nigel Marriott CStat

Independent Statistician

nigel@marriott-stats.com

On behalf of CL:AIRE at SILC Annual Forum, London, March 2020

2 Temple Street, Keynsham, Bristol, BS31 1EG, United Kingdom
Registered in England, Company Number 5577275, VAT number GB883304029